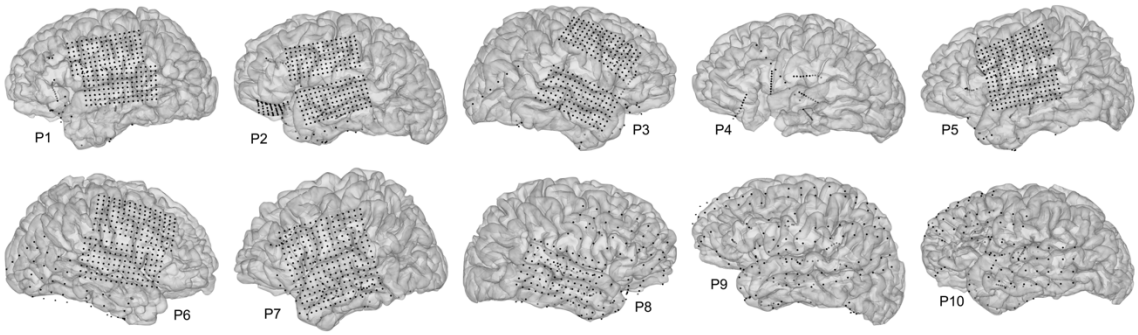

Fig. S1. ECoG coverage for all participants. Electrode placement is shown as black dots for the implanted hemisphere (semi-transparent) in the native pre-surgical MRI reconstruction space for each participant.

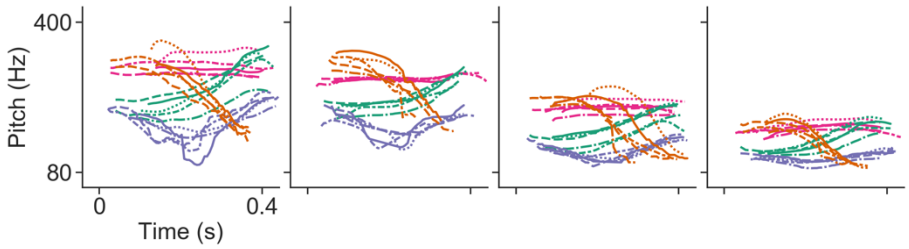

Fig. S2. Pitch patterns for all stimuli. Extracted pitch trajectories are shown for all speakers. First two cells show two female speakers. The last two cells show two male speakers. Line colors correspond to tone identity, while line styles correspond to syllable identity (see legends **Fig. 1**).

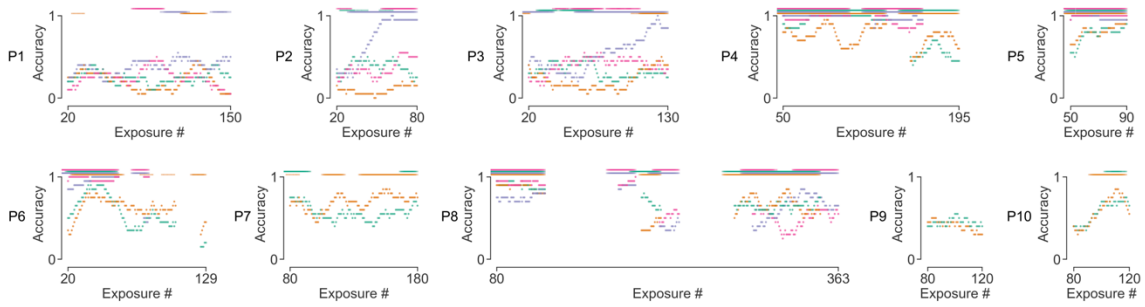

Fig. S3. Behavioral performance for all participants. Proportion of correct responses (y-axis) is shown against exposure number for each tone (x-axis). Only active training trials are shown. Horizontal lines on the top of each panel corresponds to trial windows with above-chance accuracy (FDR-P < 0.05; permutation testing). Minimum exposure number for this display item is 20, which corresponds to the size of the moving window.

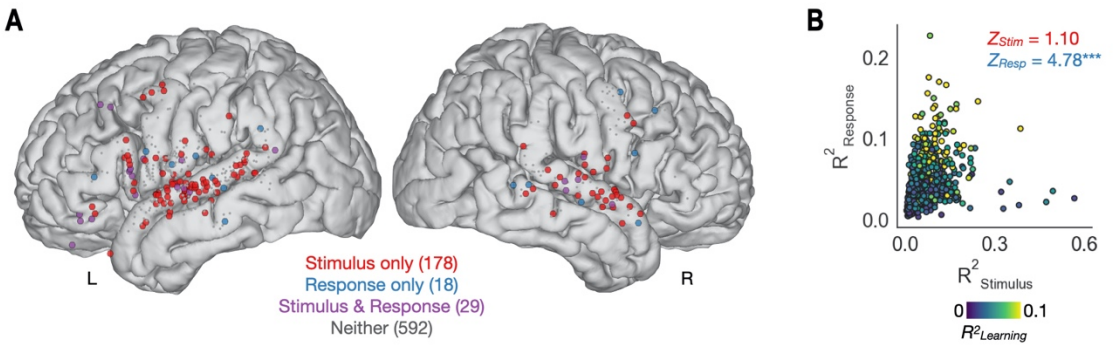

Fig. S4. Learning-associated changes modulate the neural code for stimulus-evoked percepts. Active training data shown in Fig. 2 was re-analyzed to dissociate effects of stimulus presentation and participant behavioral response, the latter of which serves as a proxy for subjective perceptual experience. The following encoding model was used: HGAt ~ (Intercept) + Syllable + Speaker + Tone + Response. (**A**) All electrodes that were significant for either stimulus or response are shown on an MNI brain. All three subtypes (red: stimulus only; blue: response only; purple: stimulus and response) are distributed across the perisylvian regions, including the superior temporal gyrus as well as the inferior frontal gyrus. (**B**) Across all sound-sensitive electrodes from all trained participants, the main behavioral effect (Tone:Accuracy:Exposure interaction; color scale) was stronger among electrodes with higher unique variance explained by behavioral response (y-axis), compared to that by stimulus identity (x-axis).

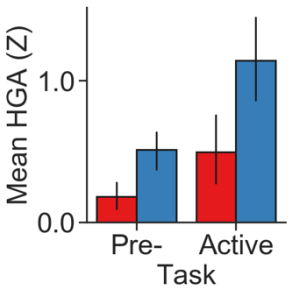

Fig. S5. Mean HGA comparison between pre-training passive and active training recording blocks. Data are shown separately for learning (red) and exposure (blue) electrodes. Errorbars correspond to SEM.

Table S1. Clinical and demographic details for participants. F: Female; M: Male; L: Left; R: Right; A: Ambidextrous.

ID	Age	Sex	Handedness	Hemisphere	Seizure Foci
P1	28	M	R	L	L Ant Temporal Lobe
P2	36	M	R	L	L Ant Temporal Lobe
P3	20	M	R	R	R Mid Amygdala and Hippocampus
P4	31	F		L	
P5	36	M	R	L	L Med Temporal Lobe
P6	27	F	A	R	R Suboccipital, Inf Temporal Gyrus, and Hippocampus
P7	49	F	R	L	L Temporal Lobe
P8	22	F	R	R	R Temporal Lobe
P9	59	F	R	L	L Med Temporal Lobe
P10	19	M	R	L	L Lat Frontal Lobe